

Preface

In our Web-connected world, a business enterprise must have timely information in order to survive. But the applications that collect and manage the information may have been developed independently, over years, using different products and technology. How can such information—scattered across multiple databases and applications—be collected and integrated for access in real time?

This issue of the *IBM Systems Journal* provides some answers to this question. An introductory essay and 10 papers discuss aspects of information integration, from research challenges to technology and products. We are indebted to A. Jhingran, H. Pirahesh, and M. Wirth, IBM Research Division, and N. Matto, IBM Software Group, for their efforts in obtaining and coordinating these papers. In addition, a paper on transaction compensation and a proposal for managing “spam” are included in the issue.

The first two papers set the stage, presenting a broad perspective on information integration. In the introductory essay, Jhingran, Matto, and Pirahesh approach the topic along three dimensions. For each dimension—data types, federation, and intelligence—they discuss current activity and research directions. In the second paper, Roth et al. provide a historical view, tracing the progression of business applications and database systems from the 1970s through the present. Four business scenarios illustrate current requirements for information integration.

In the third paper, Haas, Lin, and Roth explore the federation dimension of information integration. In database federation, middleware consisting of a relational database management system provides uniform access to a number of heterogeneous data sources. The system shields its users from the need

to know anything about the sources, including how the data in the sources are modeled and managed. The benefits are demonstrated through a number of usage scenarios.

The next three papers focus on access to integrated information. The first, by Chamberlin, introduces XQuery, a new query language for Extensible Markup Language (XML) data sources. The author describes the XQuery requirements, which were defined by the World Wide Web Consortium, and provides a tutorial on the language and examples of its use.

The next two papers, both by Funderburk et al., continue the discussion of XQuery. The first describes XTABLES, a middleware system that provides the ability to create and query XML views of relational data. In addition, it allows XML documents to be stored and queried as relational data. Thus an XQuery request handled by XTABLES can provide unified access to both XML documents and relational data. The second paper describes an integrated database architecture that supports XQuery as well as XML extensions to SQL (Structured Query Language). The architecture allows a seamless flow, from relational data to XML and back again.

The paper by Malaika et al. approaches information integration from the perspective of the World Wide Web. The authors show how DB2* and Web services can be used together, integrating information from multiple Web service providers and exposing the collective information through Web services.

In the next paper, Somani, Choi, and Kleewein explore two important trends: the increase in kinds of information needed and in the breadth of informa-

tion distribution. Using a brokerage scenario, the authors show the need for a system that can provide access to a collection, or federation, of heterogeneous and widely distributed information sources.

Two papers discuss data analysis in real time, or on-line analytical processing (OLAP). In the first, Cody et al. propose a blend of business intelligence with knowledge management that makes use of analysis techniques from both fields. Two tools developed to explore this approach are described. In the second paper, Colossi, Malloy, and Reinwald define metadata extensions to a relational database (DB2). These extensions allow applications to retrieve multidimensional data at varying levels of aggregation, greatly improving OLAP efficiency.

The last paper on information integration gives a flow technology perspective. The authors, Leymann and Roller, describe information integration in terms of data integration and function integration. Function integration, based on flow technology and adapter technology, adds powerful capabilities to information integration, according to the authors, and they discuss various ways to couple a database system and a flow engine.

In a paper based on work supported by the IBM Faculty Program, Chessell et al. describe how existing mechanisms for transaction compensation can be extended. These extensions are described in a business processing modeling language, which packages the primary behavior and the compensation behavior together in a “compensation pair.” The extensions are described in the context of IBM’s Business Process Beans (BPBeans) technology, which is a feature of IBM’s WebSphere*.

Fahlman, in a Technical Forum article, proposes a way to handle unwanted e-mail messages and phone calls. If the senders of unsolicited messages required payment for “interrupt rights,” advertisers and fundraisers would have a strong incentive to reduce the unsolicited traffic and carefully target their audience.

The next issue of the *Journal* contains papers on autonomous computing.

Marilyn L. Bates
Associate Editor

John J. Ritsko
Editor-in-Chief

*Trademark or registered trademark of International Business Machines Corporation.